

questions
de communication

Questions de communication

14 | 2008

Moteurs de recherche. Usages et enjeux

Vie privée et quête de la perfection des moteurs de recherche audiovisuels

Privacy and the Drive for the Perfect Audiovisual Search Engine

Michael Zimmer

Traducteur : Anne Fages et Brigitte Simonnot



Édition électronique

URL : <http://journals.openedition.org/questionsdecommunication/841>

DOI : 10.4000/questionsdecommunication.841

ISSN : 2259-8901

Éditeur

Presses universitaires de Lorraine

Édition imprimée

Date de publication : 1 décembre 2008

Pagination : 115-139

ISBN : 978-2-86480-981-4

ISSN : 1633-5961

Référence électronique

Michael Zimmer, « Vie privée et quête de la perfection des moteurs de recherche audiovisuels », *Questions de communication* [En ligne], 14 | 2008, mis en ligne le 16 janvier 2012, consulté le 07 mai 2019. URL : <http://journals.openedition.org/questionsdecommunication/841> ; DOI : 10.4000/questionsdecommunication.841

Tous droits réservés

MICHAEL ZIMMER

School of Information Studies

The University of Wisconsin-Milwaukee, États-Unis

zimmerm@uwm.edu

VIE PRIVÉE ET QUÊTE DE LA PERFECTION DES MOTEURS DE RECHERCHE AUDIOVISUELS

Résumé. — L'internet devient de plus en plus une plateforme de création, de stockage et de partage de contenus multimédias produits par les utilisateurs. Par conséquent, les fournisseurs de moteurs de recherche s'efforcent d'étendre les fonctionnalités et la portée de leurs technologies pour rendre ces contenus audiovisuels aussi « universellement accessibles et utiles » que leurs homologues textuels. Cependant, l'émergence d'outils de recherche audiovisuels puissants suscite de nouvelles inquiétudes quant à la façon dont les moteurs de recherche s'immiscent dans la vie privée. Cette contribution fait le point sur certaines technologies émergentes de recherche dans les données audiovisuelles comme l'intégration des images produites par les utilisateurs dans les index des moteurs, le marquage et les métadonnées enrichies des photographies, ainsi que les outils de reconnaissance faciale destinés aux utilisateurs. En outre, elle montre comment, dans le contexte de la quête de la perfection des moteurs de recherche, ces technologies sont une source de préoccupation croissante pour la protection de la vie privée.

Mots clés. — Moteurs de recherche, recherche audiovisuelle, contenus produits par les utilisateurs, web 2.0, reconnaissance faciale, vie privée, surveillance.

Les moteurs de recherche¹ Web se sont imposés en tant qu'outils omniprésents et cruciaux pour réussir à naviguer dans la sphère toujours en expansion de l'information en ligne. Le moteur de recherche le plus important au niveau mondial, Google (2005a)², est emblématique du but de tous les moteurs généralistes lorsqu'il affirme sa mission : « Organiser l'information mondiale et la rendre universellement accessible et utile ». C'est la raison pour laquelle Google Inc., comme la plupart des opérateurs d'outils de recherche, s'efforce de créer « un moteur de recherche parfait capable de traiter et d'assimiler toute l'information mondiale » (Google, 1999)³ et de comprendre « exactement ce que vous voulez dire et vous donner en retour précisément ce que vous voulez » (Google, 2007)⁴. La quête d'un tel moteur a donné naissance à des outils de recherche puissants qui permettent l'accès à des dizaines de milliards de sources d'information en ligne, traitent des millions de requêtes chaque jour et fournissent des résultats pertinents et souvent personnalisés, tout cela gratuitement.

Avec l'essor des réseaux à grande vitesse, la baisse des coûts des technologies numériques, l'omniprésence des webcams et des appareils photo intégrés aux téléphones portables, et avec le développement de sites Web innovants qui permettent le stockage de contenus et l'interaction sociale – comme Flickr ou Facebook –, le Web devient de plus en plus une plateforme de création, de stockage et de partage de contenus multimédias produits par les utilisateurs. Par conséquent, les opérateurs de moteurs de recherche s'efforcent d'étendre la puissance et la portée de leurs robots de collecte pour inclure dans leurs index des images, des contenus audio et vidéo. Ils mettent à disposition leurs propres plateformes de stockage, de recherche et de partage d'images

¹ L'auteur remercie les organisateurs et les participants du *Forum on Quaero : A Public Think Tank on the Politics of the Search Engine* [Forum sur Quaero : groupe de réflexion public sur la politique du moteur de recherche], hébergé par l'Académie Jan van Eyck et l'atelier expert *Audiovisual Search : Regulatory Challenges for Audiovisual Abundance* [Recherche audiovisuelle : défis réglementaires pour l'abondance audiovisuelle], hébergé par l'Institut du Droit de l'information de l'université d'Amsterdam, pour avoir permis la présentation de cette recherche et pour leurs remarques. L'auteur adresse également ses remerciements aux relecteurs externes de *Questions de communication* pour leurs précieuses suggestions. Ce travail a été en partie soutenu par le Projet Société de l'information de la Yale Law School (New Haven, Connecticut).

² « Organize the world's information and make it universally accessible and useful », *Company overview*, <http://www.google.com/corporate/index.html> (consulté le 03/06/06).

³ « A perfect search engine [that] will process and understand all the information in the world », « Google Receives \$25 million in equity funding [press release] », *Google Press Center*, 07/06/99, <http://www.google.com/press/pressrel/pressrelease1.html> (consulté le 18/08/06).

⁴ « Exactly what you mean and give back exactly what you want », *Our philosophy*, <http://www.google.com/intl/en/corporate/tenthings.html> (consulté le 27/03/07)

et de vidéos, et parfois se spécialisent dans les contenus exclusivement multimédias, comme le projet de moteur de recherche européen Quaero. Bref, le moteur de recherche parfait fait la part belle à l'audiovisuel.

La puissance croissante et l'omniprésence des moteurs de recherche font qu'ils n'échappent pas à la controverse : les biais qui les affectent, la censure (ou ses contournements), les violations de droit d'auteur et de marque, ainsi que la liberté d'expression en ligne sont des sujets constants de préoccupation. Toutefois, les plus importants concernent les risques que présentent ces outils de recherche puissants pour la vie privée (Sullivan, 2003 ; Associated Press, 2005 ; Mills, 2005 ; Zimmer, 2008a, 2008b). Créés dans l'optique de rendre les contenus audiovisuels aussi « universellement accessibles et utiles » que leurs homologues textuels, ces nouveaux et puissants outils de recherche audiovisuelle suscitent un regain d'inquiétude chez les universitaires et les autorités de réglementation sur ces problématiques⁵. Du fait de la nature spécifique des contenus audiovisuels, les problématiques liées à la vie privée que soulève la quête de la perfection des moteurs de recherche audiovisuels méritent une attention particulière.

Cette contribution fait le point sur les dernières tendances dans le domaine des technologies de recherche d'information audiovisuelle – l'intégration d'images produites par les utilisateurs dans les index des moteurs, le marquage et les métadonnées enrichies des photographies et les outils de reconnaissance faciale destinés aux utilisateurs – et met en évidence les risques qu'ils constituent pour la vie privée. Elle s'appuie sur une analyse qualitative et une synthèse des informations publiques sur les technologies de recherche audiovisuelle, qui proviennent des communiqués de presse officiels des compagnies et de la documentation sur les produits, mais aussi de la couverture par la presse et d'entretiens avec des dirigeants d'entreprise. Nous avons nourri la réflexion par l'analyse de diverses technologies de recherche d'information, que ce soit à partir d'un équipement standard et l'utilisation des technologies de recherche audiovisuelle, ou encore des analyseurs de protocoles du réseau pour voir ce qui se passe « derrière l'interface » des outils utilisés.

⁵ Par exemple, en septembre 2007, l'Académie Jan van Eyck (Maastricht, Pays-Bas), en collaboration avec la Maison Descartes, l'Institut français des Pays-Bas, a tenu un forum critique sur le projet européen de recherche d'information audiovisuelle Quaero (voir http://www.janvaneyck.nl/0_2_3_events_info/arc_07_quaero_thinktank.html). S'ensuivit, en avril 2008, un atelier d'experts sur les défis de réglementation auxquels doivent faire face les moteurs de recherche d'information audiovisuelle, à l'Institut pour le Droit de l'information de l'Université d'Amsterdam.

La quête de la perfection des moteurs de recherche

Depuis que les premiers moteurs de recherche ont commencé à fournir un moyen d'interfacer les contenus du Web, a débuté la quête du « moteur de recherche parfait », celui qui sera capable d'indexer toute l'information disponible et de fournir des résultats rapides et pertinents (Kushmerick, 1998 ; Andrews, 1999 ; Gussow, 1999 ; Mostafa, 2005), et qui donnerait des résultats intuitifs basés sur les recherches précédentes et l'historique général de navigation de l'utilisateur (Pitkow *et al.*, 2002 ; Teevan *et al.*, 2005). Par exemple, le moteur parfait saurait si une recherche sur l'expression « Paris Hilton » est censée aider l'utilisateur à localiser des chambres d'hôtel dans la capitale française ou à trouver les derniers potins sur la jeune célébrité, et lui fournirait uniquement les résultats adéquats dans chacun des cas. Il est clair que les entreprises qui travaillent sur les moteurs de recherche ont des motivations financières pour réussir la « recherche d'information parfaite » : offrir à l'utilisateur des résultats personnalisés contribuerait probablement à le rendre dépendant d'un moteur donné, tout en augmentant son exposition à la publicité des sites partenaires et en améliorant les chances pour que l'utilisateur recoure à des services payants. De même, les moteurs peuvent augmenter leurs tarifs publicitaires quand les annonces sont précisément présentées devant les yeux d'utilisateurs ayant des besoins et des centres d'intérêt qui y correspondent – par exemple, quelqu'un qui chercherait des renseignements touristiques plutôt que les derniers commérages sur une célébrité⁶. Pour le journaliste John Battelle (2004 ; 2005 : 228), le moteur de recherche parfait devrait fonctionner de la manière suivante :

« Imaginons qu'on puisse poser n'importe quelle question et obtenir une réponse exacte et surtout *parfaitement adaptée*. Cette réponse correspond au contexte et à l'intention de la question et sait, avec une précision surnaturelle, qui vous êtes et pourquoi vous demandez cela. Une réponse capable d'incorporer tout le savoir du monde accessible à la recherche (aux formats texte, vidéo ou audio) pour accomplir sa tâche. Une réponse capable de distinguer les interrogations directes (« Qui fut le troisième président des États-Unis ? ») des questions plus nuancées (« Dans quelles circonstances le troisième président des États-Unis abjura-t-il [...] sa conception de l'esclavage ? »). [...] Cette recherche idéale disposerait également d'une mémoire parfaite. Elle saurait ce que vous avez déjà vu et ne confondrait pas requête de découverte (procédant d'un désir d'apprendre de nouvelles choses) et envie de retrouvailles (lorsqu'on souhaite retrouver quelque chose) ».

Interrogé sur ce que serait pour lui le moteur de recherche parfait, Sergey Brin (*in* : Ferguson, 2005 : 40) de Google Inc. a répondu : « Comme

⁶ S. Hansell, « Microsoft plans to sell search ads of its own », *The New York Times* (26/09/05 : C1, C8).

l'esprit de Dieu » [« *Like the mind of God* »]. Pour atteindre un tel idéal d'omniprésence et d'omniscience, le moteur de recherche parfait doit avoir à la fois la « portée parfaite » [*perfect reach*] pour donner accès à toute l'information disponible en ligne et le « rappel parfait » [*perfect recall*]⁷ pour fournir des résultats personnalisés et pertinents qui tiennent compte des habitudes d'un chercheur d'information donné.

La portée parfaite

Pour réussir à atteindre la portée nécessaire à la réalisation de la recherche d'information 2.0, les moteurs de recherche en ligne amassent dans leurs index une quantité phénoménale d'informations sur les contenus du Web. Loin de se limiter aux pages Web en HTML, les opérateurs de moteurs ont indexé des médias très diversifiés trouvés en ligne, notamment des images, des fichiers vidéo, pdf et autres documents informatisés. Par exemple, en 2005, Yahoo prétendait indexer plus de 20 milliards d'éléments, dont plus de 19,2 milliards de documents Web, 1,6 milliards d'images et plus de 50 millions de fichiers audio et vidéo (Mayer; 2005). De plus, les moteurs numérisent de plus en plus le « monde physique », en ajoutant le contenu de livres à succès, les fonds des bibliothèques universitaires, des cartes géographiques et des images satellites à leurs index en croissance permanente. Outre leurs index extensibles et diversifiés, les moteurs d'aujourd'hui atteignent aussi la « portée parfaite » en développant divers outils et services permettant aux utilisateurs d'organiser et d'utiliser l'information dans des contextes que la recherche d'information en ligne ne prenait pas en considération. Cela comprend les plateformes de communication et de réseaux sociaux, la gestion de données financières, les achats et la recherche de produits en ligne, la gestion de fichiers informatisés et une navigation sur l'internet améliorée.

En combinant ces deux aspects de la portée parfaite – des index de recherche extensibles et divers produits d'organisation de l'information – le moteur de recherche idéal donne aux utilisateurs le pouvoir de chercher et de trouver; et les met en contact avec toutes les formes d'information dont ils ont besoin dans leur vie quotidienne. La portée du moteur de recherche parfait permet aux utilisateurs de chercher et d'accéder à pratiquement tout le contenu en ligne, mais aussi de communiquer, de naviguer, d'acheter et d'organiser leurs vies, aussi bien en ligne que hors ligne.

⁷ En recherche d'information, le rappel [*recall*] mesure la proportion de documents pertinents effectivement retrouvés par un système de recherche. Nous avons conservé la traduction par le terme de rappel même si, ici, le terme est employé dans un sens plus général de « mémoire » ou « remémoration ».

Le rappel parfait

En complément de la portée parfaite du moteur idéal, les opérateurs de moteurs souhaitent fournir le rappel parfait à chaque chercheur d'information, tout en permettant la personnalisation des services et de la publicité. Pour obtenir ce rappel, les moteurs de recherche Web doivent pouvoir identifier et comprendre les souhaits, besoins et désirs des internautes lorsqu'ils font des recherches en ligne. Par exemple, pour différencier le contexte et l'intention de la requête « Paris Hilton », le moteur parfait devrait savoir si l'internaute a récemment témoigné quelque intérêt pour un voyage en France ou s'il passe un temps significatif à rechercher en ligne des commérages sur les célébrités.

Le principal moyen de personnaliser les résultats d'une recherche est de se fier aux habitudes de recherche d'information des utilisateurs et à leur historique (Pitkow *et al.*, 2002 ; Speretta, Gauch, 2005 ; Teevan *et al.*, 2005). Pour collecter les historiques de recherche des utilisateurs, la plupart des moteurs gèrent des journaux détaillés qui enregistrent chaque requête traitée par leurs serveurs, l'adresse IP d'où provient la requête, les pages vues et les résultats sur lesquels l'internaute a cliqué (Google, 2005b)⁸. Les moteurs de recherche misent aussi beaucoup sur les *cookies* pour distinguer les utilisateurs et tracer leur activité d'une session à l'autre, et les poussent de plus en plus à créer des comptes – ce qui les aide à associer un utilisateur donné à son activité en ligne. On explique à l'utilisateur que ces informations sont recueillies dans le but d'améliorer à l'avenir son expérience de la recherche en ligne. Google, par exemple, le présente en ces termes : « Nous utilisons cette information afin d'améliorer la qualité de nos services et pour d'autres raisons commerciales »⁹, alors que le moteur Ask.com présente également ses motivations économiques comme suscitées, selon le discours de l'entreprise, par la nécessité de ce rappel parfait dans la quête de la recherche d'information parfaite : « Nous recueillons [...] des informations anonymes afin d'améliorer la qualité générale de votre expérience en ligne, notamment grâce au contrôle du produit, à son amélioration, à une publicité ciblée et à la monétisation

⁸ Google Privacy faq, http://www.google.com/privacy_faq.html (consulté le 03/03/06) ; IAC Search & Media, 13/07/05, *Privacy policy for Ask.com*, <http://sp.ask.com/en/docs/about/privacy.shtml> (consulté le 06/01/07) ; Yahoo!, 11/11/06, *Yahoo! Privacy policy*, <http://info.yahoo.com/privacy/us/yahoo/details.html> (consulté le 06/01/07).

⁹ « We collect [...] anonymous information to improve the overall quality of the online experience, including product monitoring, product improvement, targeted advertising, and monetizing commercially oriented search keywords », *Google Privacy faq*, http://www.google.com/privacy_faq.html (consulté le 03/03/06).

des mots clés des requêtes à visée commerciale »¹⁰. Étant donné la palette étendue de produits et de services offerts par les opérateurs de moteurs, les historiques de leurs serveurs peuvent contenir bien plus que les requêtes des internautes¹¹. Les requêtes sur divers contenus en ligne – images, actualités, vidéos, livres, recherches scientifiques, produits, musique, etc. – sont aussi probablement enregistrées, tout comme les liens sur lesquels ils ont cliqué et les statistiques d'usage des services qui ne sont pas liés à la recherche d'information mais que fournissent ces mêmes entreprises, tels que le courrier électronique, les groupes de discussion, les services de cartes ou d'itinéraires et ainsi de suite.

Pour résumer, la quête du moteur de recherche parfait a conduit les opérateurs de moteurs de recherche à étendre la portée de leurs index et à permettre d'interroger de nombreuses sources d'information diversifiées (à la fois en et hors ligne) ; elle a également amélioré le rappel pour certains besoins, habitudes et souhaits spécifiques des utilisateurs. Si l'on considère les étapes accomplies vers la portée et le rappel parfaits que nous venons d'évoquer, les moteurs semblent se rapprocher de plus en plus de cet idéal d'omniscience et d'omniprésence, et sont sur le point de « devenir l'instrument de pouvoir parfait du XXI^e siècle, qui voit tout, qui sait tout, qui peut tout » (Ayers, 2003 : 4).

Vie privée et moteur de recherche parfait

L'enthousiasme suscité par la recherche du moteur parfait – celui qui promet une large couverture, profondeur et pertinence – n'est pas totalement bénigne. La route qui mène à ce moteur est jonchée d'obstacles liés aux préoccupations sur les biais, la censure – ou son contournement –, la violation du droit d'auteur ou du droit des marques, et la liberté de parole (e.g. Introna, Nissenbaum, 2000 ; Elkin-Koren, 2001 ; Gasser, 2006 ; Goldman, 2006 ; Grimmelmann, à paraître). Toutefois, les inquiétudes les plus fortes sont liées aux risques potentiels d'un tel moteur de recherche idéal pour la vie privée (Sullivan 2003 ; Associated Press, 2005 ; Mills, 2005 ; Zimmer 2008a, 2008b). En effet, la portée et le rappel nécessaires à la réalisation du moteur parfait ont des implications spécifiques sur la vie privée des individus.

¹⁰ IAC Search & Media, 13/07/05, *Privacy policy for Ask.com*, <http://sp.ask.com/en/docs/about/privacy.shtml> (consulté le 06/01/07).

¹¹ Wikipedia liste plus de cent produits et services différents proposés par Google et plus de cinquante par Yahoo.

Vie privée et portée parfaite

Pour obtenir la portée parfaite, les moteurs doivent indexer autant de pages Web et de sources d'information en ligne que possible pour proposer la plus grande base de données de résultats de recherche potentiels. Parmi les milliards de pages Web indexées par les moteurs, se trouvent des pages contenant des informations personnelles sur les individus, notamment les pages personnelles (qu'elles soient actives ou abandonnées), les messages postés dans les forums de discussion, les *curriculum vitae* en ligne, les comptes rendus de réunions publiques, les registres fonciers et les archives des tribunaux. Peu sont ceux qui n'ont pas encore été touchés par le « long bras du robot de collecte de Google » explique le journaliste Neil Swidey (2003 : 11) en faisant référence au rayon d'action de l'index Web de Google :

« Il s'agissait peut être d'une farce stupide entre étudiants ou d'un message imprudent posté sur un forum Internet à l'université. Vous vous êtes peut être lancé dans une diatribe lors d'une réunion du conseil municipal ou signé une pétition sans prendre le temps de la lire. Ou peut être avez-vous vécu un divorce difficile. Vous pensiez peut-être avoir tourné la page. Et bien Google ne partage pas votre avis.

Même si votre embarrassant bagage était déjà en grande partie public, il était hors de portée de tous sauf du professionnel intrépide et vraiment curieux. À présent, aux États-Unis, dans les États où les décisions de justice sont publiées en ligne, et grâce à la facilité de la fonction en un seul clic de Google, vous pouvez lire tous les détails sordides du divorce de votre voisin sans plus d'efforts qu'il ne vous en faut pour vérifier si vous avez de nouveaux messages ».

Une recherche par vanité, c'est-à-dire une recherche en ligne sur son propre nom, peut révéler un nombre surprenant d'informations personnelles : feuilles d'imposition, documents judiciaires, certificats de mariage, actes et informations d'enregistrement sur les listes électorales par exemple. « Googler » quelqu'un avant une première rencontre est devenu une pratique courante (Lobron, 2006). Près d'un internaute sur quatre a déjà recherché des informations en ligne sur ses collègues ou ses relations d'affaire (Sharma, 2004), et les employeurs « googlisent » leurs futurs employés avant de prendre leurs décisions d'embauche (Weiss, 2006). En moins d'une heure, un journaliste a découvert une mine de détails personnels sur la vie du PDG de Google, Eric Schmidt (Mills, 2005)¹² :

¹² Ironiquement, Google a puni le cnet pour cette publication d'information personnelle sur E. Schmidt – information trouvée grâce à leur propre moteur de recherche – par un boycott d'un an de son service d'actualité. Face à la critique générale, Google a mis fin au boycott deux mois plus tard.

« Schmidt n'en dit pas beaucoup sur lui-même dans sa page personnelle. Mais il a suffi de 30 minutes passées sur le moteur Google pour découvrir que Schmidt, âgé de 50 ans, pesait environ 1,5 milliard de dollars l'an dernier. Au début de l'année, il a mis en vente près de 90 millions de dollars d'actions Google et obtenu 50 millions de dollars de plus en vendant des actions au cours des deux derniers mois, alors que le cours de l'action faisait un bon de plus de 300 \$.

Lui et sa femme Wendy vivent dans la riche ville d'Atherton en Californie où, à l'occasion d'un dîner de collecte de fonds politiques à 10 000 \$ le plat, il y a cinq ans, le candidat aux élections présidentielles Al Gore et sa femme Tipper ont dansé alors qu'Elton John chantait à plein poumons « *Bennie and the Jets* ». Schmidt a aussi parcouru le désert au festival artistique *Burning Man* au Nevada et est un pilote amateur passionné ».

L'expansion des index qui permettent la recherche d'information, aussi bien en termes de couverture que de niveau de détail, dans le but d'atteindre la « portée parfaite » a amoindri le sentiment « d'intimité par non-visibilité » [« *privacy through obscurity* »] des utilisateurs (Swidey, 2003 ; Ramasastry, 2005). Une multitude d'informations sur les individus jusque là difficiles à trouver est désormais facilement accessible grâce aux moteurs de recherche (Wright, Kalalik, 2000). Par conséquent, il est aujourd'hui de plus en plus facile d'exploiter le nombre croissant d'informations personnelles publiées en ligne pour s'engager dans des « cyberfilatures » (Tavani, Grodzinsky, 2002), recouper des données et fouiller les informations personnelles dans des bases de données qui, autrement, seraient hétérogènes (Garfinkel, 2000), et de construire des dossiers numériques sur les individus (Solove, 2004). Le spécialiste de l'éthique, Herman Tavani (2005 : 40), a notamment écrit sur la facilité avec laquelle les moteurs de recherche permettent de collecter, regrouper et analyser des informations personnelles : « Les forums de discussion contiennent des liens vers l'information postée par tel ou tel individu, ils peuvent par conséquent donner aux usagers des moteurs une bonne idée de ses centres d'intérêt et de ses activités¹³. Il semble que toute l'information actuellement disponible sur les sites Web et accessible aux moteurs n'ait pas été nécessairement placée là par les personnes concernées ni même avec leur autorisation explicite ». Il se peut qu'un individu ne sache pas que son nom figure parmi ceux répertoriés dans une ou plusieurs de ces bases de données accessibles aux moteurs, et encore moins comment fonctionnent les moteurs et leur capacité à retrouver des informations personnelles à partir de diverses sources d'information en ligne.

¹³ Une recherche sur mon nom révèle des messages (il est vrai que je les avais oubliés) dans les forums de discussion Usenet du début des années 90, sur des sujets allant du droit à l'avortement, au catholicisme, en passant par le féminisme, le marketing et le tableur Lotus 1-2-3.

C'est John Battelle (2005 : 168) qui résume le mieux l'anxiété que suscite la portée parfaite des moteurs de recherche : « Que faire quand une information légalement publique devient, disons, *trop* publique ? [...] Que se passe-t-il quand tout ce qui vous concerne – vrève mention dans la lettre d'information de votre école primaire (actuellement en ligne bien sûr) ou divagations revanchardes d'un amant éconduit – suit votre nom pour l'éternité ? ».

Vie privée et rappel parfait

La portée des moteurs de recherche se rapprochant de plus en plus de la perfection, les préoccupations quant aux risques potentiels pour la vie privée sont d'autant plus exacerbées que les moteurs de recherche sont aujourd'hui capables de recueillir et de rassembler une grande quantité d'informations personnelles et morales sur leurs utilisateurs pour atteindre cet objectif du rappel parfait. L'inquiétude la plus manifeste porte sur le suivi des requêtes individuelles et la capacité de les associer à des adresses IP spécifiques. Cette pratique de recueil et de mémorisation des données relatives aux requêtes pour atteindre le « rappel parfait » a attiré l'attention du grand public en janvier 2006, lorsqu'une source a révélé que, dans le cadre des efforts déployés par le gouvernement des États-Unis pour faire respecter une loi sur la pornographie en ligne, le Département américain de la Justice a demandé à un juge fédéral de contraindre le moteur de recherche Google à lui remettre les enregistrements de millions de requêtes de ses utilisateurs (Hafner, Richtel, 2006 ; Mintz, 2006). Google a résisté mais trois de ses concurrents, America Online (AOL), Microsoft et Yahoo se sont pliés à des assignations semblables du gouvernement de produire leurs enregistrements des requêtes (Hafner, Richtel, 2006). Plus tard, la même année, AOL rendait publiques environ 20 millions de requêtes provenant de 658 000 de ses utilisateurs pour essayer de soutenir la recherche universitaire sur l'analyse des requêtes soumises aux moteurs de recherche (Hansell, 2006). Malgré les tentatives d'AOL de rendre les données anonymes, certains particuliers restaient identifiables uniquement à partir de ces historiques de recherches, qui comportaient des termes de requêtes correspondant aux noms des utilisateurs, à des numéros de sécurité sociale, à des adresses, des numéros de téléphone et autres informations personnelles permettant l'identification (McCullagh, 2006a).

Ces affaires ont dévoilé au grand public les pratiques de mémorisation des requêtes, générant une vague d'inquiétude parmi les chercheurs d'information sur la présence d'un tel contrôle systématique de leurs activités de recherche d'information en ligne (Barbaro, Zeller Jr, 2006 ; Hansell, 2006 ; McCullagh, 2006b) et amenant la presse à enquêter

et publier des rapports sur l'information que les moteurs recueillent systématiquement sur leurs utilisateurs (Glasner, 2005 ; Ackerman, 2006). À leur tour, divers groupes de pression ont critiqué l'ampleur du suivi et du recueil des requêtes que permettent ces moteurs, souvent sans même que les utilisateurs ne le sachent vraiment (voir par exemple Electronic Frontier Foundation, 2007 ; Privacy International, 2007), alors que les autorités gouvernementales des pays européens et des États-Unis commençaient à enquêter sur les pratiques et politiques de conservation des requêtes des moteurs (Associated Press, 2007 ; Lohr, 2007).

Alors que les pratiques d'archivage des requêtes Web des internautes attiraient de plus en plus l'attention du grand public, la capacité des opérateurs de moteurs à surveiller et croiser leurs activités – réalisées via la palette de produits et de services qu'ils proposent – semble nourrir de nouvelles inquiétudes (Glasner, 2005 ; Ackerman, 2006 ; Waters, 2007). Le fait d'encourager les utilisateurs à créer des comptes, la promotion croisée des services pour empêcher les utilisateurs de quitter le site du moteur et l'usage de *cookies* persistants dans ces services, permettent aux moteurs de recherche de recueillir et de regrouper une quantité importante d'informations personnelles et morales sur leurs utilisateurs, qui va bien au-delà des quelques mots clés sur lesquels ils font des recherches : cela comprend notamment les actualités qu'ils consultent, leurs centres d'intérêt, les blogs qu'ils suivent, les livres qu'ils aiment et autres détails privés sur leur identité, leurs idéaux politiques, leur état de santé, leur vie sexuelle, leur religion, leur situation financière et leurs préférences de consommation (voir Zimmer, 2008b). Il en résulte ce que John Battelle (2003) nomme une « base de donnée des intentions » :

« Ces informations représentent, une fois regroupées, un conteneur des intentions de l'humanité, une base de données massive des souhaits, besoins, désirs et passions que l'on peut découvrir, assigner, archiver, suivre et exploiter dans toutes sortes d'objectifs. Un tel monstre n'a encore jamais existé dans l'histoire de la culture. Il a toutefois de bonnes chances de grandir de manière exponentielle à l'avenir. Cet artefact peut nous révéler des choses extraordinaires sur qui nous sommes et ce que nous voulons comme culture ».

De même que nos habitudes quotidiennes – comme l'utilisation des cartes de crédit, des distributeurs bancaires, des téléphones portables ou des systèmes de péage automatiques – laissent derrière nous d'innombrables « traces virtuelles » de nos activités, la quête du rappel parfait nécessaire à la construction du moteur idéal a pour effet d'engendrer une infrastructure robuste de surveillance des données ayant le pouvoir de contrôler et d'enregistrer l'historique des recherches d'information des utilisateurs, les messages électroniques, les billets postés sur les blogs et leurs habitudes générales de navigation, donnant « une excellente source de connaissance de ce qu'une personne *pense*, et pas seulement de ce qu'elle fait » (Hinmann, 2005 : 23).

Intégrer l'audiovisuel dans la recherche parfaite

Les images numériques et les contenus vidéo proliférant sur l'internet et le Word Wide Web, les opérateurs de moteurs ont longtemps cherché à inclure ces contenus audiovisuels dans leurs index. En 1995, Alta Vista créait l'événement en étant le premier moteur à proposer la recherche d'images, de contenus audio et vidéo, une fonctionnalité bientôt reprise par les autres principaux moteurs de l'époque, comme Lycos, HotBot et InfoSeek. Ces tout premiers services de recherche audiovisuelle intégraient l'accès à des fichiers multimédias en se basant principalement sur l'extension des noms de fichiers (.gif, .avi, .au) et les références textuelles extraites du contexte comme les noms de fichiers, les métadonnées, les liens hypertextes ou les mots environnants (Schwartz, 1998 : 979). Aujourd'hui, si certains moteurs spécialisés existent encore (Picsearch, par exemple, permet de chercher dans plus de deux milliards d'images), les services de recherche audiovisuelle sont l'apanage des grands moteurs généralistes. Google, Yahoo, Microsoft et Ask donnent tous la possibilité de rechercher des fichiers images et vidéo, en utilisant les mêmes méthodes de recherche qu'Alta Vista a développées dix ans plus tôt.

Même si l'identification du contenu des images pour la recherche d'information dépend encore largement des métadonnées, on voit surgir des innovations dans la recherche audiovisuelle, en particulier des améliorations à la fois techniques et sociales de ces services. Avec l'expansion des sites web de partage de photographies, un secteur aujourd'hui dominé par Flickr, le nombre d'images personnelles disponibles en ligne a explosé. Tout en fournissant une plateforme robuste et facile d'usage pour le téléchargement et le partage de photos en ligne, l'innovation principale de Flickr réside dans le fait qu'il permet aux utilisateurs d'étiqueter les photos avec leurs propres descripteurs. Les chercheurs d'information peuvent ainsi trouver des images concernant tel domaine comme par exemple un nom de lieu ou un sujet thématique. Dès son rachat par le moteur de recherche Yahoo, la base de données de Flickr, riche de plus de 100 millions de photographies, a rapidement été intégrée aux résultats de recherche d'image de Yahoo, augmentant de façon considérable la portée de son offre de recherche audiovisuelle¹⁴.

¹⁴ Yahoo!, « Flickr-izing image search », *Yahoo! Search Blog*, 26/06/07, <http://www.ysearchblog.com/archives/000466.html> (consulté le 10/07/07).

L'exploitation des étiquettes des photos créées par les auteurs des images dans Flickr a aidé Yahoo à améliorer la précision de son moteur de recherche d'images : plutôt que d'utiliser le texte environnant ou simplement le nom de l'image pour déterminer sa pertinence vis-à-vis de la requête, Yahoo peut à présent tirer profit des multiples étiquettes placées à bon escient par les individus pour décrire l'image. Google n'a pas tardé à développer sa propre méthode d'amélioration des résultats de recherche d'image en mettant au point son *Image Labeler* (Lensen, 2006). Les utilisateurs sont associés aléatoirement par paire et une image de la base de données de Google leur est présentée. Chaque internaute propose des mots clés pour décrire l'image, sans voir les propositions de l'autre. S'ils trouvent tous deux le même mot clé en moins de 120 secondes, ils gagnent des points. Bien que le dispositif soit divertissant pour les utilisateurs, c'est aussi une manière intelligente pour Google de s'assurer que les images sont associées aux mots clés les plus appropriés.

Parallèlement à l'utilisation croissante d'étiquettes produites par les individus et leur mise en réseau pour aider à identifier et à rechercher des images, des avancées dans la recherche d'images basées sur les technologies de reconnaissance automatique de formes ont considérablement amélioré la recherche audiovisuelle (Lew et al., 2006 ; Vasconcelos, 2007). Plutôt que de reposer sur des métadonnées créées par les individus, les systèmes fondés sur le contenu cherchent à reconnaître et cataloguer différentes formes, couleurs et textures dans les images, ce qui leur permet par exemple de distinguer l'image d'un chien brun de celle d'une pomme rouge. La capacité des systèmes de recherche d'images d'identifier automatiquement les visages et de les discerner fait depuis peu l'objet de toutes les attentions, à la fois pour des applications policières et commerciales (Zhao et al., 2003). De plus en plus, ces systèmes de reconnaissance faciale se développent dans le monde de la grande consommation et sont intégrés aux moteurs de recherche du Web. Ainsi, après avoir acheté Neven Vison, une entreprise qui développe une technologie de détection et de reconnaissance d'objets et de personnes dans les images, Google¹⁵ a-t-il amélioré son service de recherche d'image en limitant les résultats aux seules images montrant des visages. Par exemple, une recherche d'image simple sur « Zimmer » donne différents résultats : de la voiture hors série du même nom à diverses images de chambres sélectionnées au hasard (*Zimmer* signifie chambre en allemand) ainsi que plusieurs photos de personnes. Mais en sélectionnant le type de contenu « visages » dans le

¹⁵ « A better way to organize photos ? », *Official Google Blog*, 15/08/06, <http://googleblog.blogspot.com/2006/08/better-way-to-organize-photos.html> (consulté le 15/08/07).

formulaire de recherche d'image avancée, on obtient uniquement des visages de personnes.

Récemment, divers services de recherche audiovisuelle spécialisés en reconnaissance faciale ont été mis sur le marché. Lancé début 2006, le service de recherche de photo Riya propose des services de reconnaissance faciale et de reconnaissance de texte pour identifier des visages dans la collection de photographies numériques de l'utilisateur et les étiqueter en conséquence¹⁶. Une fois que l'utilisateur a entraîné manuellement le système en étiquetant le visage d'une personne sur un grand nombre de photos, le logiciel prend la relève et peut analyser, identifier et étiqueter automatiquement cette personne sur toutes les autres photos qui seront chargées ultérieurement dans l'album de l'utilisateur (Ribeiro, 2006). Un million de photos ont été téléchargées sur Riya dans les deux premiers jours de son lancement et sept millions de photos au cours des sept premières semaines (Arrington, 2006). Fort de son succès, Riya a annoncé son projet d'étendre ses technologies de reconnaissance faciale au-delà des collections personnelles de photos en les appliquant désormais aux images en ligne. Grâce à ce moteur de recherche visuelle encore en projet, les utilisateurs pourront soumettre une image à Riya (le dessin d'un tapis, une photographie de sac à main ou même le visage d'une personne) et obtenir en retour des images semblables trouvées en ligne (Arrington, 2006).

Le site suédois Polar Rose projette des innovations similaires pour la recherche d'image¹⁷. Plutôt que de compter sur les contributions des internautes qui téléchargent leurs albums d'images personnels pour qu'ils soient analysés (comme Riya), Polar Rose prévoit de mettre à disposition une extension Web (*plug in*) qui identifiera les visages dans les images sur pratiquement n'importe quel site, permettant aux utilisateurs d'étiqueter les visages qu'ils reconnaissent lors de leurs navigations sur le Web (Simonite, 2006). Ces données seront ensuite envoyées au serveur central de Polar Rose, ce qui permettra à n'importe qui, regardant une image où apparaît un visage donné, de dire de qui il s'agit, et de confirmer ou de corriger sa description. Ainsi l'internaute qui verrait une photo représentant de manière évidente Bill Gates, mais étiquetée comme représentant Steve Jobs, pourra-t-il rendre Polar Rose plus « intelligent » en corrigeant l'erreur. Les utilisateurs pourront aussi chercher en ligne d'autres photos contenant ce visage, en se fondant sur les données biométriques uniques créées par Polar Rose pour ce visage.

¹⁶ Voir <http://www.riya.com/>.

¹⁷ Voir <http://www.polarrose.com/>. À l'heure où nous rédigeons cette contribution, Polar Rose est encore en version d'essai et d'accès privé.

Pour résumer, les services de recherche audiovisuelle ont fait l'objet de nombreux progrès technologiques, notamment en intégrant des images produites par les utilisateurs dans leurs index interrogeables, grâce à l'étiquetage généralisé des fichiers images et au déploiement de systèmes de reconnaissance de visages capables d'apprendre avec l'aide de l'utilisateur. Grâce à ces innovations, les fournisseurs ont pu étendre la portée de leurs services de recherche en ligne pour intégrer une quantité toujours plus grande d'images produites par les utilisateurs et accessibles sur l'internet. En fait, certains moteurs de recherche ont pour seule vocation de donner accès à des contenus audiovisuels, comme le projet scientifique européen de recherche d'information multimédia Quaero, dont l'un des buts est de faciliter l'accès aux archives du patrimoine audiovisuel et aux bibliothèques numériques¹⁸. En même temps, alors que ces nouveaux outils de recherche audiovisuelle donnent aux utilisateurs l'opportunité de chercher des visages, des sons, des vidéos et des lieux, les opérateurs de moteurs peuvent mieux accéder aux souhaits et désirs des utilisateurs, améliorant leur capacité à mieux « se souvenir » de leurs intentions. Par exemple, si le moteur sait qu'un utilisateur recherche essentiellement des vidéos de célébrités plutôt que des photos de sites touristiques européens, cela peut l'aider à fournir des résultats plus pertinents pour une recherche sur les mots « Paris Hilton ». En somme, le moteur de recherche parfait devient de plus en plus audiovisuel.

Vie privée et moteur audiovisuel parfait

L'amélioration de la portée et du rappel dans la quête du moteur parfait fait peser des menaces spécifiques sur la vie privée. Le fait que le moteur idéal devienne de plus en plus audiovisuel fait émerger de nouvelles menaces peut être encore plus pesantes. Les tentatives d'ajout de composants audiovisuels au moteur de recherche idéal exacerbent les menaces pour la vie privée, à la fois dans la recherche de l'amélioration du rappel et de la portée parfaites.

Vie privée et portée parfaite de la recherche audiovisuelle

Comme nous l'avons décrit, la portée parfaite des moteurs de recherche a mis à mal « l'intimité par non-visibilité » des individus en rendant visibles les sites web, les documents et autres fichiers en ligne qui étaient auparavant difficiles à localiser ou auxquels on ne pouvait pas

¹⁸ Voir <http://www.quaero.fr>

facilement accéder. L'émergence de composants audiovisuels accélère la désintégration de la vie privée par non-visibilité au fur et à mesure que les utilisateurs perdent chaque jour un peu plus le contrôle sur les possibilités d'accéder et d'identifier leur apparence en ligne. Les sites de partage d'images comme Flickr ont simplifié les fonctions de chargement, d'étiquetage et de partage des images, qu'il s'agisse des siennes propres ou de celles des autres. Avec l'ubiquité des appareils photographiques intégrés aux téléphones portables, il devient possible de filmer presque n'importe quel moment de la vie de quelqu'un, de l'étiqueter et de le mettre en ligne à la disposition de tous. Depuis l'intégration des images de Flickr aux résultats de la recherche d'images sur Yahoo, les internautes n'ont plus besoin d'être membres de la communauté Flickr (ou même de savoir qu'elle existe) pour accéder aux images que ce service stocke. Par conséquent, n'importe qui pourra être pris en photo, laquelle sera ensuite mise en ligne, étiquetée avec son nom et découverte grâce au moteur de recherche Yahoo sans même qu'il le sache¹⁹. Comme il arrive souvent que des personnes soient prises en photo (notamment dans les lieux publics), cette nouvelle possibilité d'avoir des images étiquetées et accessibles *via* la recherche à des millions de personnes menace les normes actuelles de vie privée du fait que l'apparence de tout un chacun est mise à la disposition du grand public.

Cette inquiétude est exacerbée par la prédominance de métadonnées détaillées. Les appareils photo numériques utilisent généralement le format d'échange de fichier Exif, qui enregistre diverses métadonnées dans chaque image, y compris la date et l'heure, les paramètres de réglage de l'appareil, une image réduite de prévisualisation sur l'écran de l'appareil photo ou dans un logiciel, et éventuellement des informations concernant les droits d'auteur. De nombreuses images sont géocodées à l'aide d'étiquettes d'information sur le lieu, soit manuellement par l'utilisateur, soit automatiquement grâce à un système de GPS intégré à l'appareil photo lui-même. Ces métadonnées détaillées peuvent être consultées (et souvent on peut faire des recherches dessus) sur des sites de partage de photographie en ligne, ce qui permet aux utilisateurs de savoir précisément quand et où telle photo a été prise. Par conséquent, de plus en plus de photos de personnes sont disponibles en ligne à leur insu et, par ailleurs, elles sont accompagnées d'une mine d'informations détaillées sur le lieu et le moment où cette personne se trouvait quand la photo a été prise, détériorant encore un peu plus la toute relative « intimité par non-visibilité ».

¹⁹ Par exemple, une recherche sur « Michael Zimmer » *via* la recherche d'image de Yahoo montre une photo de moi donnant une conférence, étiquetée avec mon nom et postée sur Flickr; tout cela à mon insu.

L'émergence d'outils robustes de reconnaissance faciale pour identifier et rechercher des images en ligne donne un autre exemple où la puissance de la portée du moteur de recherche audiovisuel parfait menace la vie privée (voir, pour une approche globale, Granick, 2005 ; Inman, 2007). Plus besoin des utilisateurs pour identifier le visage d'une personne dans une photo Flickr et l'étiqueter : il est désormais facile de découvrir des visages à partir de ses caractéristiques biométriques uniques. Avec des services tels que Riya ou Polar Rose, un utilisateur peut charger des douzaines de photos personnelles pour entraîner le logiciel à apprendre à reconnaître son propre visage, ce qui rendra plus facile l'étiquetage de futures images. Cependant, Riya exploite ce profil facial pour permettre également aux autres utilisateurs de rechercher et d'identifier des images. Ainsi un utilisateur différent pourra-t-il soumettre une photo trouvée au hasard en ligne ; le logiciel Riya analysera sa base pour apparier les visages, et étiquettera automatiquement tout visage reconnaissable parmi les nouvelles images. Les utilisateurs pourront aussi balayer et analyser tout le Web pour des visages qu'ils ont entraîné Riya à reconnaître. Un autre utilisateur avait auparavant peu de moyens à disposition pour identifier une personne sur une photo mais, avec les outils de reconnaissance faciale comme Riya et Polar Rose, l'identification de visages inconnus devient bien plus facile. En outre, Polar Rose facilite la recherche en ligne d'autres images comportant le visage de cette personne. Cela peut amener à trouver des images sur des sites de réseaux sociaux, des pages personnelles, des sites Web professionnels, etc. Même si toutes ces images étaient visibles publiquement, il aurait été difficile pour un utilisateur commun de savoir qu'elles existent toutes, ou de savoir où les trouver. Avec ces nouveaux outils sociaux de reconnaissance faciale, les internautes perdent leur capacité de contrôler le contexte et la communauté dans laquelle certaines images peuvent être visualisées.

Vie privée et rappel parfait de la recherche audiovisuelle

Les menaces sur la vie privée qu'amène avec elle l'extension de la portée de la recherche audiovisuelle ne sont pas les seules : des menaces similaires se profilent avec l'amélioration du rappel du moteur de recherche idéal grâce à la présence de contenus audiovisuels. À l'instar des requêtes générales des internautes qui sont suivies à la trace et enregistrées par les moteurs de recherche pour aider à prévoir leurs souhaits et leurs besoins, les termes des requêtes audiovisuelles – des noms, des lieux, de la pornographie, etc. – peuvent également être interceptées et potentiellement associées aux comptes utilisateurs. Plus que simplement embarrassantes, certaines requêtes pourraient éclairer les points de vue sur la société ou la politique d'un individu, et éventuellement s'avérer d'une

légalité douteuse. Tout comme les représentants de la loi ont manifesté leur intérêt pour le pistage ou le recueil des activités générales de recherche, ils pourraient bien prêter attention aux recherches audiovisuelles sur des éléments controversés ou illégaux comme la pornographie enfantine, les caricatures du prophète Mahomet, les souvenirs nazis ou les contenus protégés par un droit d'auteur. Bref, l'ajout de contenus audiovisuels aux services de recherche du Web rend possible des requêtes spécifiques qui menacent de rendre la « base de données des intentions » plus détaillée et potentiellement plus préjudiciable.

Renégocier le défi faustien

En s'efforçant de rendre les contenus audiovisuels aussi « universellement accessibles et utiles » que leurs homologues textuels, les fournisseurs de services de recherche sur le Web ont ajouté à leur offre de nouveaux outils puissants, tels que l'insertion d'images produites par les utilisateurs, un étiquetage des photographies et des métadonnées robustes, et l'émergence d'outils de reconnaissance faciale sophistiqués mis entre les mains de l'internaute ordinaire. Ces améliorations audiovisuelles sont porteuses de menaces importantes pour la vie privée, puisque la portée comme le rappel du moteur de recherche parfait incluent désormais des visages de personnes, des identifiants biométriques et l'historique des types de contenus multimédias qu'ils recherchent. Il en ressort donc une sorte de pacte avec le diable : les technologies de recherche audiovisuelle promettent de mettre au point le moteur de recherche parfait en rendant disponible une mine de contenus multimédias – dont la plupart sont produits par les utilisateurs – tout en puisant dans les métadonnées et l'idéologie Web 2.0 qui pousse les individus à identifier et marquer les images disséminées sur le Web pour faciliter les recherches d'information futures. Mais elles suscitent aussi des inquiétudes quant à la protection de la vie privée, parmi lesquelles la perte continue de « l'intimité par non-visibilité », le fait que les utilisateurs n'aient plus le contrôle sur le téléchargement d'images qui les représentent, sur l'étiquetage de leurs visages et le fait que les activités de recherche audiovisuelle puissent alimenter une « base de données d'intentions » toujours plus vaste, collectées par les opérateurs de moteurs de recherche.

De quelles options disposons-nous pour renégocier notre pacte avec le diable pour ces outils de recherche audiovisuels émergents ? D'abord, nous devons élargir notre conception de l'intimité et de la vie privée aux contenus disponibles en ligne, y compris les images mises en ligne par les particuliers ou prises dans des lieux publics. Il devient essentiel d'adopter une notion plus contextuelle de la vie privée (Nissenbaum, 2004 ; Barth *et al.* 2006 ; Zimmer, 2008a) qui admette que ce n'est pas

parce qu'une personne se trouve dans un lieu public qu'on a le droit de la photographier; de télécharger cette image sur le Web, de l'identifier; de l'étiqueter et de faire éventuellement une analyse biométrique du visage créé pour la partager avec tous. Certaines images sont destinées à un contexte particulier et cumuler toutes ces actions – même dans le but d'améliorer le moteur de recherche parfait – pourrait encore constituer une violation de la vie privée.

En second lieu, nous devons explorer de nouveaux cadres légaux et de normalisation pour faire face au flux croissant d'images personnelles en ligne et au développement de technologies sophistiquées pour les identifier, les étiqueter et permettre de les rechercher. Dans certaines juridictions où les lois actuelles restreignent la possibilité de collecter et de publier des informations personnelles sans consentement explicite, il a été envisagé d'étendre ces protections aux images en ligne qui contiennent des visages d'individus identifiables (e.g le Commissariat à la protection de la vie privée du Canada, 2008). En réaction à de telles contraintes légales (existantes ou potentielles), certains fournisseurs de services de recherche de contenus audiovisuels ont commencé à masquer les images identifiables pour tenter de protéger la vie privée des individus qui auraient pu être filmés ou pris en photo (BBC News, 2008 ; Bergstein, 2008).

D'autres mesures juridiques de protection pourraient prendre la forme du processus d'« avertissement et retrait » [*« notice and takedown »*] utilisé pour les violations présumées des droits d'auteur dans la loi américaine sur le droit d'auteur numérique (*Digital Millennium Copyright Act* ou *DMCA*). D'après le *DMCA* (et la loi sur la limite de responsabilité en cas de violation du droit d'auteur en ligne), les fournisseurs de services en ligne se voient offrir le bénéfice du « *safe harbor* » (sphère de sécurité) s'ils retirent immédiatement le contenu après que quelqu'un se soit plaint d'une violation de ses droits d'auteur. Pour faciliter cela, de nombreux site internet ont créé des procédures d'« avertissement et retrait » fiables permettant aux détenteurs de droits d'auteurs de les aviser en cas de réclamation pour violation. Par exemple, le moteur de recherche Google a mis en place des instructions et des procédures très complètes pour gérer les notifications de violation de droits d'auteur²⁰. Ces procédures pourraient être modifiées afin de prendre en compte les préoccupations liées à la protection de la vie privée des utilisateurs plutôt que celles liées à la violation des droits d'auteur. Plus précisément, plutôt qu'un détenteur de droits d'auteurs n'adresse une notification en cas d'une violation potentielle, notification suivie du retrait du contenu

²⁰ Voir <http://www.google.com/dmca.html>

en question, les particuliers devraient pouvoir envoyer un avertissement aux fournisseurs de services de recherche audiovisuelle si telle image ou vidéo constitue une violation de leur vie privée, obligeant ainsi le fournisseur à prendre des mesures raisonnables pour retirer ou masquer les informations personnelles identifiables.

Conclusion

Finalement, il existe diverses options technologiques pour atténuer les risques que présentent les nouveaux outils de recherche audiovisuelle pour la vie privée. Même en l'absence d'obligations légales ou réglementaires, les fournisseurs d'outils de recherche audiovisuelle pourraient proposer des interfaces conviviales pour les personnes souhaitant demander le retrait de leur image publiée en ligne. Par exemple, la fonctionnalité *Street View* du service cartographique de Google, qui fournit des images des rues prises au niveau du sol, offre la possibilité de signaler des images « inappropriées », notamment celles qui montrent des visages, des façades de maison et des plaques d'immatriculation. Des fonctionnalités similaires pourraient être ajoutées aux autres systèmes de recherche audiovisuelle pour permettre aux particuliers de demander de manière simple et pratique le retrait de leur image.

Une seconde option technologique relève des mêmes avancées de reconnaissance faciale qui permettent aux systèmes de recherche audiovisuelle comme Riya de fonctionner, tout en donnant à l'individu la pleine possession de la signature numérique de son visage créée par les algorithmes de reconnaissance faciale. Doté de la signature numérique de son propre visage, l'utilisateur pourrait alors disposer d'outils qui parcourent automatiquement le Web à la recherche d'images correspondant à cette signature, aidant ainsi à identifier les images mises en ligne sans son accord. Ou encore, à l'instar des services de gestion de crédits à la consommation qui aident à prévenir les usurpations d'identité en signalant automatiquement aux individus lorsque de nouvelles opérations ou activités apparaissent sur leur relevé de transactions personnel, des tiers de confiance pourraient prévenir automatiquement les utilisateurs de la mise en ligne ou de l'étiquetage d'une image correspondant à la signature numérique de leur visage, ce qui permettrait à l'utilisateur d'entreprendre l'action qu'il juge nécessaire.

Quant au fait que les activités de recherche des utilisateurs puissent être suivies et intégrées dans une « base de données des intentions », les fournisseurs de recherche audiovisuelle devraient suivre le « code de bonne conduite en matière d'information » (« *Fair Information Practice Principles* ») des commissions fédérales du commerce des États-Unis et

donner accès aux utilisateurs aux données recueillies sur leurs activités de recherche²¹. Les utilisateurs devraient bénéficier d'un accès aux données collectées et conservées à leur sujet – c'est-à-dire consulter les données enregistrées dans les fichiers d'un organisme –, et pouvoir contester l'exactitude et la complétude des données. Pour revêtir tout son sens, cette mesure devrait comporter un droit d'accès sans délai et peu coûteux aux données, un moyen simple de contester les données inexactes ou incomplètes, et de supprimer les données incorrectes ou non souhaitées des historiques de la recherche²².

En somme, le développement d'outils de recherche audiovisuelle puissants peut faire considérablement avancer la quête du moteur de recherche parfait. Mais ce n'est qu'à la condition que soient abordées les multiples préoccupations concernant la vie privée – en élargissant notre conception de la vie privée, en créant des politiques de protection et en s'engageant dans une conception technologique respectueuse de la vie privée – que nous éviterons de suivre le même chemin que Faust, en pactisant avec le diable en échange de l'accès à la connaissance universelle.

Traduction : Anne Fages (traductrice, Vendres) et Brigitte Simonnot
(université Paul Verlaine-Metz, Centre de recherche sur les médiations)

Références

- Ackerman E, 2006, « What do Google, Yahoo, AOL and Microsoft's MSN know about you ? », *San Jose Mercury News*, 19/08/06.
- Andrews P, 1999, « The search for the perfect search engine », *The Seattle Times*, p. E1, 07/02/99.
- Arrington M., 2006, « Riya 2.0 on the way ; major strategy shift », *TechCrunch*, 15/06/06, <http://www.techcrunch.com/2006/06/15/riya-20-on-the-way-major-strategy-shift/> (consulté le 11/07/07).
- Associated Press, 2005, « Google Growth yields privacy fear », *Wired.com*, 17/07/05, <http://www.wired.com/politics/security/news/2005/07/68235> (consulté le 31/03/07).

²¹ Voir la section 3, « Accès/participation » du Code de bonne conduite en matière d'information de la Commission fédérale du commerce des États-Unis, disponible à : <http://www.ftc.gov/reports/privacy3/fairinfo.shtm>.

²² Pour un nom d'utilisateur ou une adresse ip, les fournisseurs de services de recherche pourraient fournir une liste de toutes les requêtes réalisées sur une certaine période, ainsi qu'une interface utilisateur afin de modifier ou d'effacer les enregistrements dans le fichier des données mémorisées. Toutefois, l'authentification de l'identité de l'utilisateur qui demande à consulter/modifier l'historique de la recherche reste un défi majeur pour cette fonctionnalité.

- 2007, « Eu data privacy officers launch investigation into Google's internet search engine », *International Herald Tribune*, <http://www.ihf.com/articles/ap/2007/05/25/business/EU-FIN-EU-Google-Privacy-Probe.php> (consulté le 28/07/07).
- Ayers C., 2003, « Google : Could this be the new God in the machine ? », *The Times*, 01/11/03, p. 4.
- Barbaro M., Zeller Jr T., 2006, « A face is exposed for AOL searcher no. 4417749 », *The New York Times*, 09/08/06, p. A1.
- Barth A., Datta A., Mitchell J. C., Nissenbaum, H., 2006, « Privacy and contextual integrity : Framework and applications », pp. 184-198, in : s&p, *IEEE Symposium on Security and Privacy*.
- Battelle J., 2003, « The database of intentions » *Searchblog*, 13/11/03, <http://battellemedia.com/archives/000063.php>
- 2004, « Perfect search » *Searchblog*, 08/09/04, <http://battellemedia.com/archives/000878.php> (consulté le 16/05/06).
- 2005, *La révolution Google : comment les moteurs de recherche ont réinventé notre économie et notre culture*, trad. de l'anglais par D. Rueff, S. Blondeel et J.-M. Thomas, Paris, Eyrolles, 2006.
- BBC News, 2008, « Google defends street view tool », *BBC News*, 07/07/08, <http://news.bbc.co.uk/2/hi/technology/7492844.stm> (consulté le 11/07/08).
- Bergstein B., 2008, « Google begins to blur faces in street view », *USA Today*, 14/05/08, http://www.usatoday.com/tech/webguide/2008-05-14-google-street-view-blur_N.htm (consulté le 11/07/08).
- Commission à la protection de la vie privée du Canada, 2008, *Rapport annuel au Parlement 2007 : Rapport sur la Loi sur la protection des renseignements personnels et les documents électroniques*, http://www.privcom.gc.ca/information/ar/200708/2007_pipeda_f.asp (consulté le 11/07/08).
- Electronic Frontier Foundation, 2007, *Privacy and search engines*, <http://www.eff.org/Privacy/search/> (consulté le 28/06/07).
- Elkin-Koren N., 2001, « Let the crawlers crawl : On virtual gatekeepers and the right to exclude indexing », *University of Dayton Law Review*, 26, pp. 180-209.
- Ferguson C., 2005, « What's next for Google ? », *Technology Review*, 108(1), pp. 38-46.
- Garfinkel S., 2000, *Database nation: The death of privacy in the 21st century* (1st ed.), Sebastopol, CA : O'Reilly.
- Gasser U., 2006, « Regulating search engines : Taking stock and looking ahead », *Yale Journal of Law & Technology*, 9, pp. 124-157.
- Glasner J., 2005, « What search sites know about you », *Wired News*, 05/04/05, <http://www.wired.com/news/privacy/0,1848,67062,00.html> (consulté le 02/08/06).
- Goldman E., 2006, « Search engine bias and the demise of search engine utopianism », *Yale Journal of Law & Technology*, pp. 188-200.

- Granick J., 2005, « Face it: Privacy is endangered », *Wired*, 05/07/05, <http://www.wired.com/politics/law/commentary/circuitcourt/2005/12/69771> (consulté le 08/04/08).
- Grimmelmann J., à paraître, « The structure of search engine law », *Iowa Law Review*.
- Gussow D., 1999, « In search of », *St. Petersburg Times*, 04/10/99, p. 13.
- Hafner K., Richtel M., 2006, « Google resists u.s. Subpoena of search data », *The New York Times*, 20/01/06, pp. A1, C4.
- Hansell S., 2005, « Microsoft plans to sell search ads of its own », *The New York Times*, 26/09/05, pp. C1, C8.
- 2006, « AOL removes search data on vast group of Web users », *The New York Times*, 08/08/06, p. C4.
- Hinman L., 2005, « Esse est indicato in Google : Ethical and political issues in search engines », *International Review of Information Ethics*, 3, pp. 19-25.
- Inman M., 2007, « Face recognition for online photo searches sparks privacy fears », *National Geographic News*, 05/01/07, <http://news.nationalgeographic.com/news/2007/01/070105-photo-search.html> (consulté le 05/04/08).
- Introna L., Nissenbaum H., 2000, « Shaping the Web : Why the politics of search engines matters », *The Information Society*, 16(3), pp. 169-185.
- Kushmerick N., 1998, « The search engineers », *The Irish Times*, 23/02/98, p. 10.
- Lenssen P., 2006, « Google image labeler ». *Google Blogoscoped*, 01/09/06, <http://blogoscoped.com/archive/2006-09-01-n44.html> (consulté le 1/07/07).
- Lew M., Sebe N., Djeraba C., Jain R., 2006, « Content-based multimedia information retrieval : State of the art and challenges », *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2(1), pp. 1-19.
- Lobron A., 2006, « Googling your Friday-night date may or may not be snooping, but it won't let you peek inside any souls », *The Boston Globe Magazine*, 05/02/06, p. 42.
- Lohr S., 2007, « Google Deal said to bring u.s. Scrutiny », *The New York Times*, 29/05/07, <http://www.nytimes.com/2007/05/29/technology/29antitrust.html> (consulté le 27/07/07).
- Mayer T., 2005, « Our blog is growing up – and so has our index », *Yahoo! Search Blog*, 08/08/05 <http://www.ysearchblog.com/archives/000172.html> (consulté le 25/11/06).
- McCullagh D., 2006a, August 7). « Aol's disturbing glimpse into users' lives », *CNET News.com*, 07/08/06, http://news.com.com/AOLs+disturbing+glimpse+into+users+lives/2100-1030_3-6103098.html?tag=st.num (consulté le 03/12/06).
- 2006b, « Police blotter: Google searches nab wireless hacker », *CNET News.com*, 20/12/06, http://www.news.com/2100-1030_3-6144962.html (consulté le 13/06/06).
- Mills E., 2005, « Google balances privacy, reach », *CnET News.com*, 03/08/05, http://news.com.com/Google+balances+privacy,+reach/2100-1032_3-5787483.html (consulté le 07/01/07).

- Mintz H., 2006, « Feds after Google data : Records sought in u.s. Quest to revive porn law », *San Jose Mercury News*, 16/01/06, <http://www.siliconvalley.com/mld/siliconvalley/13657386.htm> (consulté le 19/01/06).
- Mostafa J., 2005, « Seeking better Web searches », *Scientific American.com*, 24/01/05, http://www.sciam.com/print_version.cfm?articleID=0006304A-37F4-11E8-B7F483414B7F0000 (consulté le 30/01/05).
- Nissenbaum H., 2004, « Privacy as contextual integrity », *Washington Law Review*, 79(1), pp. 119-157.
- Pitkow J., Schütze H., Cass T., Turnbull D., Edmonds A., Adar E., 2002, « Personalized search », *Communications of the ACM*, 45(9), pp. 50-55.
- Privacy International, 2007, « A race to the bottom : Privacy ranking of internet service companies », 09/06/07, <http://www.privacyinternational.org/article.shtml?cmd%5B347%5D=x-347-553961> (consulté le 10/07/07).
- Ramasastry A., 2005, « Can we stop zabasearch - and similar personal information search engines ? When data democratization verges on privacy invasion », *FindLaw*, 12/05/05, <http://writ.news.findlaw.com/ramasastry/20050512.html> (consulté le 12/06/06).
- Ribeiro J., 2006, « Riya photo search site readies for launch », *PCWorld*, 19/01/06, <http://www.pcworld.com/article/id,124407-c,searchengines/article.html> (consulté le 11/07/07).
- Schwartz C., 1998, « Web search engines », *Journal of the American Society for Information Science*, 49(11), pp. 973-982.
- Sharma D., 2004, « Is your boss Googling you? », *CNET News.com*, 21/10/04, http://news.com.com/Is+your+boss+Googling+you/2100-1038_3-5421210.html (consulté le 06/01/07).
- Simonite T., 2006, « Face-hunting software will scour Web for targets », *New Scientist*, 19/12/06, <http://technology.newscientist.com/article/dn10828-facehunting-software-will-scour-web-for-targets.html> (consulté le 11/07/07).
- Solove D.; 2004, *The digital person : Technology and privacy in the information age* (Ex machina), New York: New York University Press.
- Speretta M., Gauch S., 2005, « Personalized search based on user search histories », pp. 622-628, in : *Proceedings of the IEEE/WIC/ACM International conference on Web Intelligence (WI'05)*, Université de Technologie de Compiègne, 19-22 sept.
- Sullivan D., 2003, « Search privacy at Google & other search engines », *SearchEngineWatch*, 2/04/03, <http://searchenginewatch.com/showPage.html?page=2189531> (consulté le 31/03/07).
- Swidey N., 2003, « A nation of voyeurs : How the internet search engine Google is changing what we can find out about one another – and raising questions about whether we should », *The Boston Globe Sunday Magazine*, 02/02/03, p. 10.
- Tavani H. T., 2005, « Search engines, personal information and the problem of privacy in public », *International Review of Information Ethics*, 3, pp. 39-45.
- Tavani H. T., Grodzinsky F., 2002, « Cyberstalking, personal privacy, and moral responsibility », *Ethics and Information Technology*, 4(2), pp. 123-132.

- Teevan J., Dumais S. T., Horvitz E., 2005, « Personalizing search via automated analysis of interests and activities » pp. 449-456, in : *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*.
- Vasconcelos N., 2007, « From pixels to semantic spaces : Advances in content-based image retrieval », *Computer*, 40(7), pp. 20-26.
- Waters D., 2007, « What search engines know about us ». *BBC News*, 31/05/07, <http://news.bbc.co.uk/2/hi/technology/6700997.stm> (consulté le 2/04/08).
- Weiss P., 2006, « What a tangled Web we weave: Being googled can jeopardize your job search », *New York Daily News*, 19/03/06 (consulté le 7/01/07).
- Wright M., Kakalik J., 2000, « The erosion of privacy », pp. 162-170, in : R. M. Baird, R. M. Ramsower, S. E. Rosenbaum, eds, *Cyberethics : Social and moral issues in the computer age*, Amherst, N.Y, Prometheus Books.
- Zhao W., Chellappa R., Phillips P., Rosenfeld A., 2003, « Face recognition : A literature survey », *ACM Computing Surveys*, 35(4), pp. 399-458.
- Zimmer M., 2008a, « Privacy on planet Google : Using the theory of "contextual integrity" to clarify the privacy threats of google's quest for the perfect search engine », *Journal of Business & Technology Law*, 3(1), pp. 109-126.
- 2008b, « The gaze of the perfect search engine : Google as an infrastructure of dataveillance » pp. 77-99, in : Spink A., Zimmer M., eds, *Web searching : Multidisciplinary perspectives*, Dordrecht, The Netherlands, Springer.